Applied Ergonomics 1981, **12.1**, 39–45

Experiments with bar graph process supervision displays on VDUs

L.H.J.M. Verhagen

Twente University of Technology, CT Department/Ergonomics Working Group, Enschede, Netherlands
Now at N.V. Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands

Bar graph representations of process variables are compared with alternatives (stroke-type and 'T'-type, combined bar and stroke) in experiments with human subjects, using either an automatic slide projector or a closed-circuit TV system. The stroke-type appears to give superior results when used for detection of off-normal conditions.

Introduction

Computerised Visual Display Units (VDUs) are rapidly being introduced for process supervision and control. These devices allow greater freedom in structuring information. Common formats are: flow diagrams, tables, graphic symbols and bar graphs.

Bar graphs are introduced as an analogue display, sometimes resembling conventional indicator instruments. They are mainly applied for two different purposes:

a. The presentation of past values of one process variable, see Fig. 1a. This presentation resembles a normal time-history graph, see Fig. 1b. Schutz (1961) has already found that for such purposes the line graph presentation is superior to the bar graph type.
b. The presentation of present values of a series of variables in an analogue form, see Fig. 2.

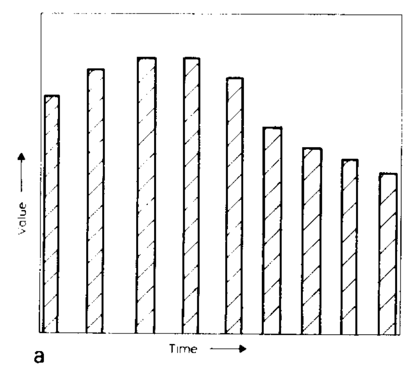

Fig. 1(a) Time history in bargraph form

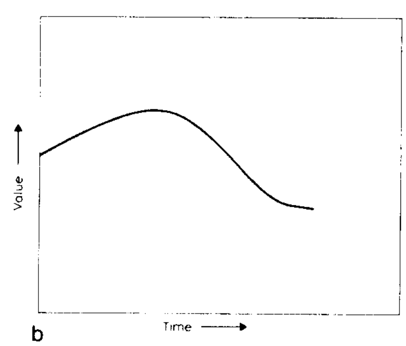

Fig. 1(b) Time history in graph form

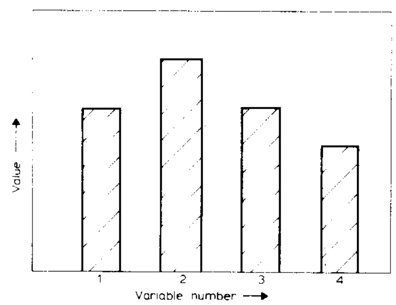

Fig. 2 Present values of variables in bar graph form

The second type of bar graph is widely used since it resembles a row of analogue indicators. The scaling of the variables, however, is quite important. Conventional analogue instruments each have an individual scale. In VDUs one sometimes sees the same principle applied: for instance eight analogue bars each with a scale on one side appear on the screen. However, there is a tendency to increase the number of variables in one display format in order to obtain a compact process overview display. In some applications, over 200 variables are represented.

It is fairly obvious that with so many variables individual scaling has to be replaced by group scaling. Scaling can be done with respect to maximum and minimum values, (perhaps with a percentage scale), set points (= midpoint on the scale) or high and low alarm limits. Colour coding and/or blink coding may be used to indicate overflow or underflow alarms.

When there is no separate alarm coding and where many process-overview pictures with blinking *and* colour changes are used, some form of 'alarm inflation' may arise (Kortlandt and Kragt, 1978). Some ergonomists speak of 'Christmas tree effects'. In the same way in situations where many alarm messages appear on one screen, single messages can easily be overlooked. In order to investigate strictly the spatial lay-out of bar graphs, we have decided to restrict ourselves to monochrome pictures without blink coding. Figs. 3a–d depict the chosen alternatives. Figs. 3a and 3b show two alternatives for a *bar graph* format, vertical and horizontal bars. Fig. 3c (stroke-type) shows an alternative to bar graphs. Only the endpoints of the bars are presented.

Finally, Fig. 3d shows an intermediate format, called T-type.

For the experiments reported here, we have asked ourselves the following questions:

a. How much time does it take to identify an over or under-flow alarm if no extra coding, such as colour change or blinking, is used?
b. How many errors are made?
c. What are the effects of different formats?
d. How many single variables can effectively be overviewed by a human operator?

In our experiments, the question of comparison of different formats was more extensively investigated. The questions on search time and errors can only be answered in conjunction with the chosen alternatives. An answer to the question of number of variables cannot be given, since the number of variables was only varied in two levels in different experiments.

Experimental method

Experiments

Four experiments were carried out. The comparisons under consideration were:

Experiments 1 and 2:	bar graph presentation (bar-type) *vs* end of bar graph only (stroke-type), 24 variables, see Figs. 4 and 5.
Experiment 3:	horizontal *vs* vertical bars, 24 variables, see Figs. 4 and 6.
Experiment 4:	T-type *vs* stroke-type, 60 variables, see Figs. 7 and 8.

With respect to the information content of the displays, the following prerequisites were made:

1. All variables would be aligned in the same direction, either horizontally or vertically.
2. All variables would be scaled between 0 and 100%.
3. The operator would either know the precise bias or scaling of the variables, or that this was irrelevant for his/her task.
4. Alarm or notification regions would be defined as above 60% and below 40% of full scale.
5. Variables would be labelled by their position on the screen. Groups would be labelled A, B, C, etc; variables within groups labelled by numbers. Of course, tag numbers or code names could have been applied too.
6. The operator's task was to inspect the screen for those variables that were within notification regions (0–40% or 60–100%).

Subjects

All subjects were young students (18–25 years). In experiments 1, 3 and 4, students from Twente University of Technology were used, in experiment 2, technical students training to be process operators were used. No subjects participated in more than one experiment. The subjects had no experience in process control tasks. All instructions were in written form with some explanation and training in the experimental room, but only to a level of understanding the task, not of some degree of experience.

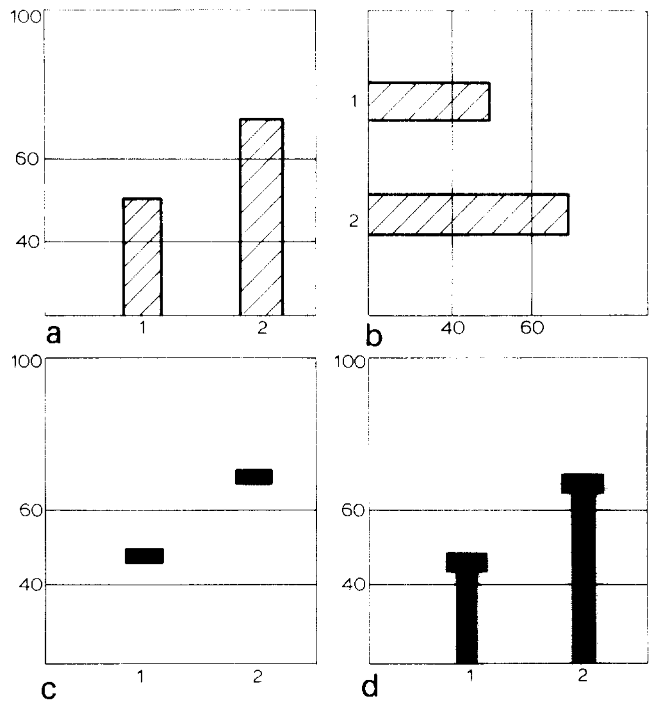

Fig. 3 Alternatives for bar graph process supervision displays (a) vertical bar, (b) horizontal bar, (c) stroke type, (d) T-type

General

All experiments were done in a quiet room. Communication with the experimenter was via an intercom. In all experiments, the process overview slides were projected with white information on a black background as is usually the case with VDUs. The experiments cannot be compared with each other in all respects since a few variations in measurement methods and tasks have been used.

In Experiment 1 (13 subjects), the time measured was that needed by subjects for counting the number of deviating variables. The time measurement was done manually. The subjects' display consisted of a transparent screen on which slides were backprojected. The subjects had to call out the answer after they had finished counting.

In Experiment 2 (12 subjects), the same apparatus with back-projection on a screen was used as in Experiment 1. However, in this case the task was to notice the deviating variables and remember their code number. The presentation time was fixed at 1 or 2 s. After the slide had disappeared, the subject had to mention the names of all deviating variables noticed. The performance measure was the percentage of all missed deviations. Another variable was introduced in this experiment. Six of the subjects were asked to start their search from a fixation point in the middle of the screen. The other six were free to choose their search strategy.

In Experiment 3 (12 subjects), slides with horizontal *vs* vertical bars were presented on a monitor via a closed-circuit TV system. Both counting and naming tasks were used. Time was measured from the appearance of the slide on the monitor to the start of the answer. The slide was not removed until the answers had been given.

A distraction task was provided by means of a binary choice generator (see also Verhagen and Lenior, 1980). On both sides of the TV monitor a small lamp was mounted. At regular intervals one of the two lamps was randomly lit. A pedal corresponding to this lamp was then to be actuated by the subject. The frequencies used were 0, 20 and 40 choices per minute. It was hypothesised that the process of regular decision-making would interfere with the identification task.

In Experiment 4 (8 subjects), finally, time was again measured, but subjects also had to call out the code number of any deviating variable immediately after it had been detected. In this experiment too the slides were presented on a TV monitor. As in Experiment 3 the influence of a binary choice task with 30 and 60 choices per minute was included as an experimental condition.

In all experiments, the conditions were balanced across subjects. Data analysis was principally done within subjects.

Results

Experiment 1

(See also Beerlage and Bonnes, 1977; Verhagen, 1977). See Figs. 4 and 5. The overall mean time needed for counting the number of deviating variables was 7·5 s for bar-type and 6·1 s for stroke-type presentation (see Table 1). An unpaired t-test (one-sided) on the results yielded significance beyond $p = 0{\cdot}005$ ($t(12) = 7{\cdot}49$).

The number of deviations that had to be found had a strong influence on search time. An analysis was made for the cases of two and nine deviations per slide. A 2 x 2 analysis of variance with type (bar and stroke) and number (2 and 9) as variables showed a significant number effect ($F(1{\cdot}48) = 30{\cdot}4$, $p < 0{\cdot}005$) but strangely enough, no type effect ($F(1{\cdot}48) < 1$). Further an interaction effect was found ($F(1{\cdot}48) = 6{\cdot}4$, $p < 0{\cdot}025$), which indicated an advantage for the bar type in the case of nine deviations. Paired t-tests showed an advantage for the stroke type in the two deviations case ($t(12) = 0{\cdot}7$, $p < 0{\cdot}01$) and no difference in the 9-deviations case.

Error analysis for Experiment 1

With the bar-type nearly twice as many errors were made as with the stroke-type (22% *vs* 12%). Many errors were made with the bar type if the number of deviations was large (12). Learning effects have not been found in the data after an initial training session.

From Experiment 1 a strong advantage appeared for the stroke type, not only in terms of search time but certainly also in terms of errors.

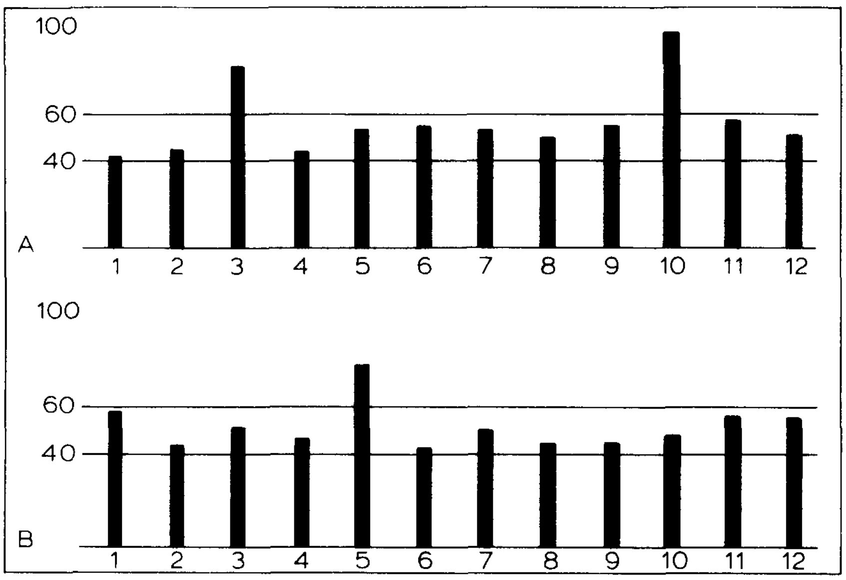

Fig. 4 Vertical bar graph for 24 variables

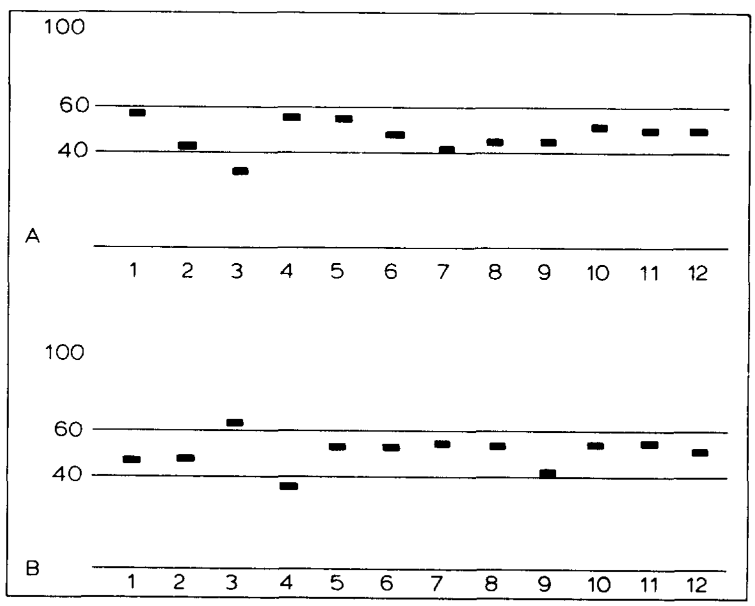

Fig. 5 Stroke graph for 24 variables

Experiment 2

(See also Steenbergen and Verhagen, 1978). See Figs. 4 and 5. Fixed presentation time was used in this experiment. The total error percentage (ie, deviations not mentioned) was 50%. The mean error scores per condition are given in Table 2.

The score of each subject consisted of the error rate for 10 slides (4x0, 2x1, 3x3, 1x6 deviations), so one score for each of the four conditions is shown in Table 2. No significant differences in error rate were found between the two search modes (starting from centre fixation point or free strategy). A paired t-test on the differences between bar and stroke showed significant results ($t(11) = 6{\cdot}74$, $p < 0{\cdot}0005$) in favour of the stroke-type. The six deviation cases did not influence the relative results strongly. If one leaves out the errors made in the six deviation slides, the results remain unchanged.

With the stroke-type more errors were made in naming the deviating variables. This means that the deviation was noticed, but given the wrong name. Most commonly an error of one place was made, eg, A2 instead of A3, as in Fig. 5. With the stroke-type, this error was made in one third of the cases (11·5 out of 37·8%) and with the bar-type in about one twentieth of the cases (2·3 out of 51·5%). Apparently, the bar-type makes it easier to determine the name which is at its bottom. In total (both types), more errors were made in detecting low alarms than high alarms (48% *vs* 36·2%) and more errors in the lower row B than in the upper row A (51·5% *vs* 35·3%). The latter result is not surprising if one assumes that subjects start scanning at the upper left corner and scan first row A, then row B.

Also it seemed that with the stroke-type more low alarms were missed than with the bar-type, but these results were not significant. More specifically, five out of 12 subjects had more low alarm errors with the stroke-type, five subjects with bar-type and two had equal scores.

In conclusion, for Experiment 2, one may state that the stroke-type could be scanned more completely than the bar-type in the same period of time. However, when under the quite extreme time stress of one or two seconds there is a danger that, with the stroke type, errors are made in finding the corresponding name of the variable. The result that more low alarms were missed than high alarms with both bar- and stroke-types needs further experimental investigation.

Table 1: Mean search time for counting of deviating variables and errors in Experiment 1

	Bar	Stroke
Mean search time	7·5 s	6·1 s
Proportion of errors	0·22	0·12

Table 2: Results of Experiment 2. Proportion of errors (ie, missed deviations) with fixed presentation time

Presentation time	Bar	Stroke	
1 s	0·61	0·48	proportion of errors
2 s	0·42	0·27	

Experiment 3

(See also Gredt, Maaskant and Trouwer, 1979). See Figs. 4 and 6. A comparison between horizontal and vertical bars was made in this experiment and both the tasks of Experiment 1 (counting) and Experiment 2, (naming of deviating variables) were used. The performance measure was the time needed for the task. The subjects had to remember the result until they found the full answer. The data of this experiment have been analysed with sign tests (Siegel, 1956). Matched pairs were taken within subjects.

With the counting task, a comparison was made between horizontal and vertical bars. The counting time was shorter for vertical than for horizontal bars with identical slides (85 times were shorter, 42 times were longer ($p < 0{\cdot}0002$)).

The two tasks, counting and identification, were compared in the case of slides where there was no deviation. One would expect the search times to be about the same,

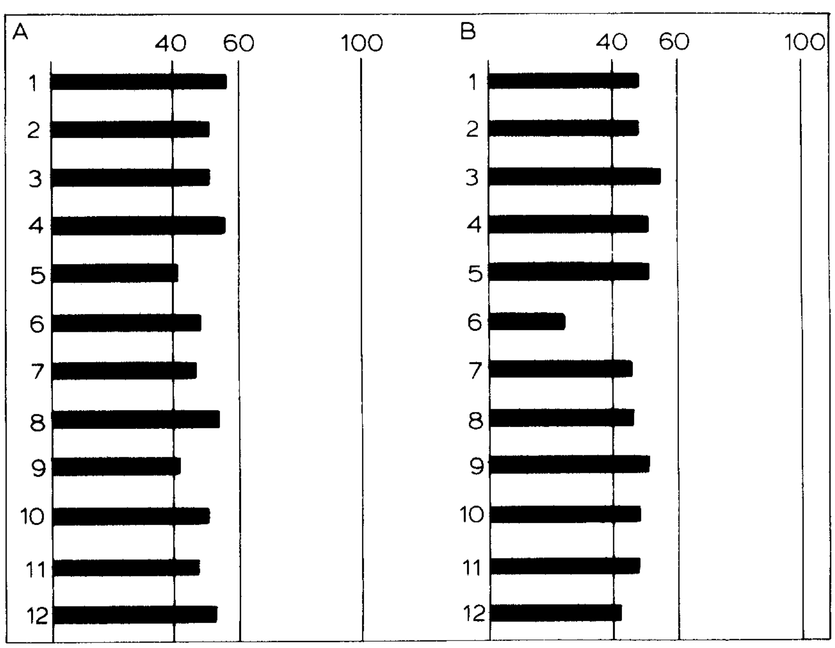

Fig. 6 Horizontal bar graph for 24 variables

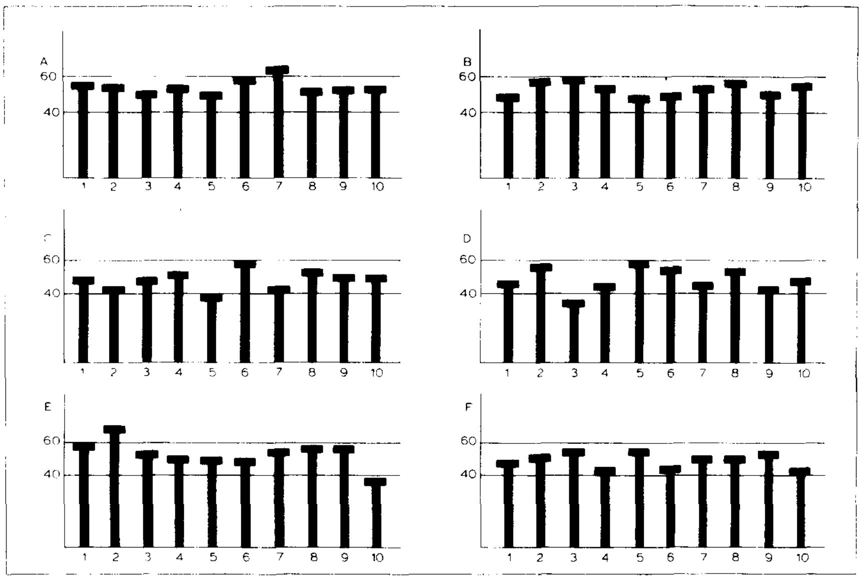

Fig. 7 Vertical T-graph for 60 variables

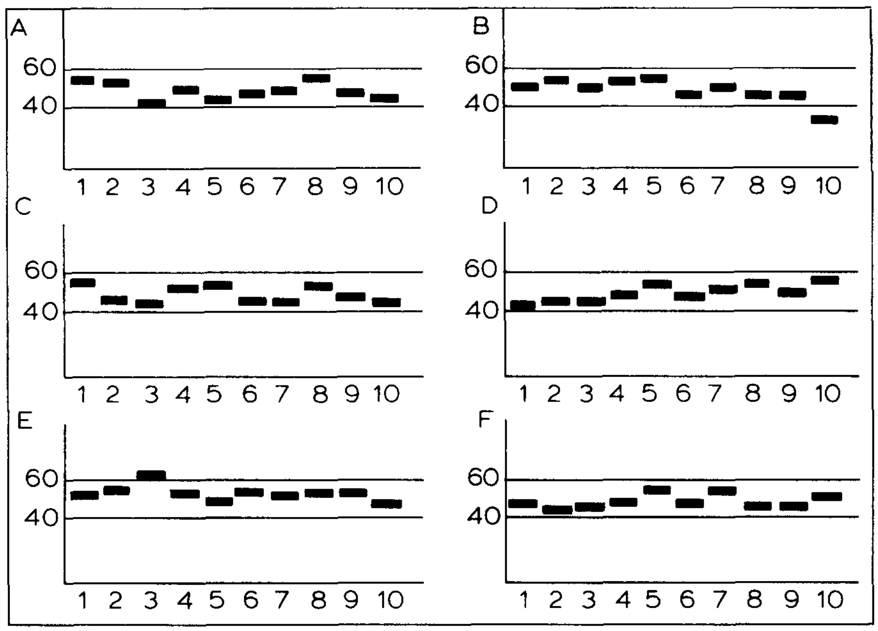

Fig. 8 Stroke graph for 60 variables

since no different cognitive action is needed. Surprisingly, the search times for the counting task were shorter (6·7 s) than for the identification task (7·4 s) (17 times were shorter, seven times were longer, $p < 0{\cdot}04$). However, with a parametric t-test no significance was found.

Together with the identification task, a distraction task was included in the form of a binary choice task with 0, 20 and 40 choices per minute. For all three distraction task conditions no significant differences between horizontal and vertical bar-type were found. Neither were significant differences found in the case of tests performed on all deviation conditions separately (0, 1, 2, 3, 4 and 6 deviations).

There were, however, significant results in the mutual comparison of three distraction tasks. With 20 choices, search times were longer than with no choices; and with 40 choices, search times were longer than with 20 choices. All significance levels were beyond 0·05.

In conclusion one may say that no preference may be expressed for horizontal or vertical bars. Only for the relatively simple counting task, a quick overview, do the vertical bars seem somewhat better.

With the identification task, code numbers of deviating variables had to be memorised until all deviations were found. Apparently this memorising process is strongly affected by a binary choice task.

Experiment 4

(See also Rademaker and Schulte, 1979). See Figs. 7 and 8. In this experiment stroke-type versus T-type with 60 variables and two levels of binary choice task (30 and 60 choices per minute), were used. Because of the strong influence of a binary choice task on the memorising process of code numbers, it was decided that in this experiment the subjects had to call out the code name of a deviating variable as soon as it was noticed. In the experiments, slides with 0, 1, 2, 3, 4, 5, and 6 deviations were present. The overall results for mean search times are presented in Table 3.

The differences between choice task levels (ie, 30 and 60) are not significant. This result is fairly surprising since most subjects found the 60 choices condition very difficult. The differences between stroke-type and T-type are significant (for the 30 choice case $t(7) = 5{\cdot}35$, $p < 0{\cdot}001$, for the 60 choice case: $t(7) = 2{\cdot}6$, $p < 0{\cdot}025$). About the same level of significance is reached if one analyses the data for each case of deviations separately (0 to 4 only). If the number of deviations increases, search times increase as well, which is not surprising (see Table 4).

In this experiment, special attention was paid to the results of Experiment 2, relating to the detection of low alarms against the detection of high alarms. In Experiment 2, there was a tendency for low alarms to be less well detected than high alarms.

In Experiment 4, slides were prepared with all low or all high alarms with one, two and three deviating bars. In this way, search times for all low alarms could be compared with search times for high alarms.

In Table 5, the overall results with respect to search times are presented. The two distraction task conditions are combined and erroneous answers ignored.

Both for stroke-type and for T-type the difference in search times is significant (paired t-tests, p is 0·05). It is surprising that for the stroke-type, low alarms are detected more quickly than high alarms. With the T-type the result is as expected.

Looking more closely at the data, it appeared, however, that the result for the stroke-type can be fully attributed to the three deviations case (9·5 *vs* 11·8 s) and for the T-type to the 1 deviation case (11·2 *vs* 10·5 s). Hence it is difficult to reach clear-cut conclusions on this matter.

Error analysis for Experiment 4

All subjects made more errors with the T-type than with the stroke-type. Total error score for the stroke-type was 8·4%, for the T-type 16·6%. This result is comparable with that of Experiment 1 (12 *vs* 22%) both in an absolute and in a relative sense.

Many errors were made during the first session in Experiment 4, especially if the subject started with the 60 choices condition. In total, more errors (omissions) were made in low alarms than in high alarms, for all sessions together, as well as for later sessions only.

Since all conditions were balanced across subjects it was necessary to check whether any learning effects were present. Rank orders of mean search times per subject and per session showed that the average order is fully random (tested with Kendalls coefficient of concordance test, W = 0·02, not significant).

Conclusions

1. An attractive alternative for bar graph process supervision displays on VDUs seems to exist in the form of a stroke-type of display. (Experiments 1, 2 and 4.)
2. With respect to the alarm detection tasks as described, there is no difference in performance between horizontal and vertical bars. (Experiment 3.)
3. The memorising of code numbers is very much influenced by distraction. (Experiment 3.)
4. Task complexity influences results of experiments as described in this paper. In Experiment 3, with counting tasks, a difference in results appeared in favour of the vertical bar. This difference disappeared with the more complicated identification task.

Table 3: Results of Experiment 4. Mean search times for S- and T-type presentation and two levels of binary choice task

Type of display	Distraction task (choices/min) 30	60	
S-type	11·0	10·9	(mean search
T-type	13·5	13·1	time in seconds)

Table 4: Results of Experiment 4. Mean search times for S- and T-type presentation by number of deviations

Type of display	Number of deviations 0	1	2	3	4	
S-type	7·7	8.9	10·2	11·9	12·8	(mean search time
T-type	9·9	10·7	12·5	13·6	15·5	in seconds)

Table 5: Results of Experiment 4. Search times of all low alarm display versus all high alarm displays

Type of display	Alarm level low alarm	high alarm	
S-type	10·0	10·8	(mean search time
T-type	12·3	11·9	in seconds)

5. With bar (or T-) graphs, nearly twice as many alarm detection errors are made compared with stroke type graphs. (Experiments 1 and 4.)
6. Within stroke type presentation, most errors are misinterpretations of, at most, one place. Within bar type presentation most errors take the form of missed alarms. Low alarms are especially easily overlooked. (Experiments 2 and 4.)

Discussion

The questions we asked in section 1, are, of course, not fully answered. A few remarks, however, can be made.

Question (a) in section 1 referred to the time needed to identify over and underflow alarms. 100% right answers took about 6 s with the stroke-type presentation of 24 variables in Experiment 1. Experiment 2 showed that a 50% score was yielded in only 1 s.

Question (b) referred to the number of errors made. Experiments 1 and 4, different with respect to both subjects and experimental method, revealed a surprisingly high error score of about 20% for bar or T-type presentation, mainly because of missed low alarms. The stroke-type produced about 10% of errors. This seems still fairly high if one takes into account that the subjects should take their time in order to reach 100% accuracy.

Question (c) has been answered most clearly. Stroke-type is superior to bar graph type. A possible explanation for this result is that the bar itself distracts the operator from his/her main task: detection of high and low alarms.

In many experiments on search tasks, display density has been found to increase search time (see references in Penn, 1975). This result is found both if the display becomes denser with meaningful information *and* with meaningless information (so called noise).

Question (d) was on the maximum number of variables that can effectively be overviewed in one format. Of course, these experiments are not suited to answer this question. There is, however, the influence of number of variables on search time. For instance, with stroke-type, the average search time for 24 variables was 6 s, for 60 variables 12 s. It seems not unreasonable to assert that under time stress, the number of errors will be very high if the number of variables is high (eg, larger than 100).

Apart from these questions, another interesting point arises. The memorising of code names should be brought to a minimum in process supervision tasks.

Experiment 3 showed a strong influence of a relatively simple distraction task like the binary choice task. A solution to this problem may be that the operator presses a function key whose position on the keyboard corresponds to the displayed position of the deviating variable. In this way, the computer system memorises the code names for the operator.

Of course, many questions can still be asked on this topic. Can a stroke-type presentation be laid out like the horizontal bar version? Does performance (with respect to both time and error) change if alarm cues such as blinking and colour change are introduced, or if many formats like these are to be supervised by one operator?

However, the general principle of the ergonomics approach still remains true. First determine as precisely as possible the task that has to be done with a specific format. If the task resembles those presented here, then a stroke-type, less dense display format may be chosen as an attractive alternative to bar graphs.

Acknowledgement

The author wishes to thank his former colleagues at Twente University, T.M.J. Lenior and J.E. Rijnsdorp, for their support and criticism during the preparation of this paper.

References

Beerlage, H., and Bonnes, H.
1977 The influence of two representation formats on meter reading (Dutch). Unpublished experiment report, CT Dept, THT.

Gredt, M., Maaskant, H., and Trouwer, J.
1979 Horizontal and vertical bar-graphs compared (Dutch). Unpublished experiment report, CT Dept, THT.

Kortlandt, D., and Kragt, H.
1978 *Journal A,* **19**, 3. Ergonomics and the struggle against alarm inflation.

Penn, R.F.
1975 Human Factors in the design of computer displays: a review. Loughborough University of Technology, HUSAT Memo No 109.

Rademaker, S., and Schulte, M.
1979 Comparison of S-type and T-type display formats (Dutch). Unpublished experiment report, CT Dept, THT.

Schutz, H.G.
1961 *Human Factors,* **3**, 7. An evaluation of formats for graphic trend displays.

Siegel, S.
1956 Non-parametric statistics for the behavioural sciences. McGraw-Hill, Tokyo.

Steenbergen, A., and Verhagen, L.
1978 Results of an experiment (Dutch). Unpublished experiment report, CT Dept, THT.

Verhagen, L.H.J.M.
1977 Preliminary experiments for the comparison of proposed VDU pictures for process control. Internal Note, Working Group Ergonomics, THT.

Verhagen, L.H.J.M., and Lenior, T.M.J.
1980 Search time experiments with distraction tasks as a tool to display format design. In preparation for submission to *Human Factors.*